

An Enterprise Level Metadata Service

J. Steven Hughes

Cynthia Wong

Daniel Crichton

Jim U'Ren

Xpace 2001

XML for Space Data

ESOC, Darmstadt, Germany

5th/6th December 2001

OUTLINE

- **The Problem**
- **The Key to a Solution**
- **A Legacy Example**
- **An Enterprise Level Metadata Service**
 - Metadata Registry
 - Resource Registry
- **Future Work**

The Problem

- **NASA Chief Information Office
Information Technology Proposal Excerpt**
 - Advances in Internet and distributed technologies have made interoperability across enterprise architectures a reality. However, NASA, with its vast data resources across the Centers, is in need of an enterprise data architecture that enables data integration and exchange services at institutional and Agency levels.

The Problem (cont)

- **The many custom-built data systems that support past and current missions present several problems to users either trying to use the data directly or develop systems to access the data**
 - There exist few common terms across data systems even for the most common concepts.
 - There is no standard language or protocol for the interchange of metadata.
 - There is no Agency-wide registration site for metadata collections .
 - There is no system to support the location and identification of metadata for reuse.
 - There is no overall and few discipline-specific control authorities for managing metadata standards to maintain consistency.
 - No widely accessible metadata resources exist to support location and navigation techniques.
 - As data volume, types, and complexity increase, metadata requirements for locating and describing the data increase.
 - Metadata captures the information known about data. Knowledge management is dependent on metadata for navigation, interoperability of distributed systems, and collaboration for communities of practice. Data mining is dependent on metadata for successful searching and knowledge discovery.

The Key to a Solution

- **Design a metadata service for a Agency-wide data architecture.**
 - Metadata is widely accepted as a major component to interoperability since it provides the vocabulary with which the enterprise creates a language for machine-to-machine and human-to-machine interaction
- **Design a metadata service that can be deployed either locally within a domain or globally across domains.**
 - Will function as an active repository for registered metadata resources.
 - Will manage and organize data elements within an ISO/IEC 11179 data registry context.
 - Will create structured sets of concepts and terms
 - Make these available for indexing and retrieval of registry content.
 - Determine relationship between existing NASA terminology standards and ISO/IEC 11179
 - Adopt XML as a common for metadata capture and exchange

Legacy Data Dictionaries

1. NASA's Office of Space Science

1. Planetary Science
2. Astrophysics
3. Space Physics (Sun Earth Connection)

2. Data Dictionaries mean different things to different people

1. **Vocabularies** – human readable collections of terms and definitions pertaining to a domain
2. **Data element dictionaries** – machine interpretable collections of data elements
3. **Schemas** (information models) – structured, machine interpretable collections of information models consisting of structured relationships between data elements.

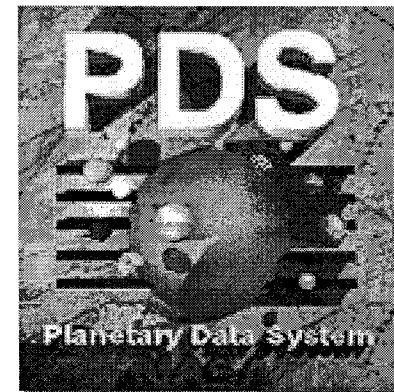
3. Dictionaries do not communicate with each other.

The Planetary Data System

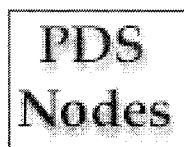
The PDS manages, preserves, and disseminates the large volume of unique and valuable data returned by Solar System Exploration missions

Key PDS Products:

- High quality, peer-reviewed data archives
- Value-added data products
- Educational data products
- Science expertise for researchers
- Data distribution to planetary scientists
- Interface to active missions and mission planning



<http://pds.jpl.nasa.gov>



New Mexico
State Univ.



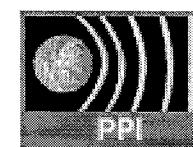
Wash Univ.
St. Louis



JPL/USGS
Flagstaff



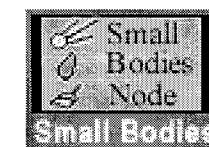
JPL



UCLA



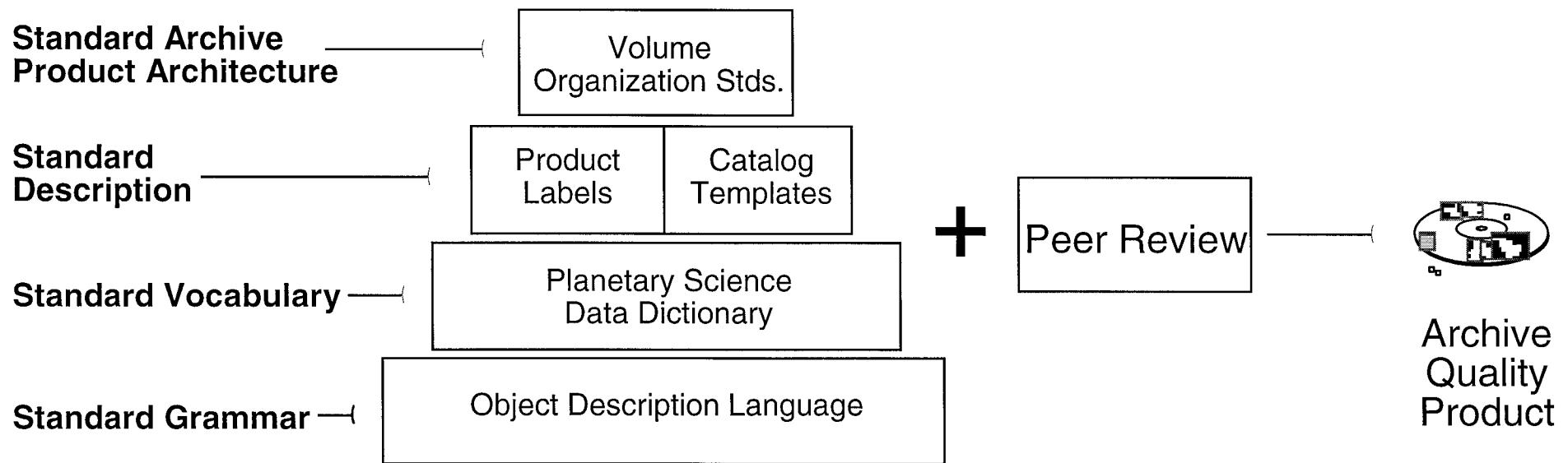
NASA
Ames



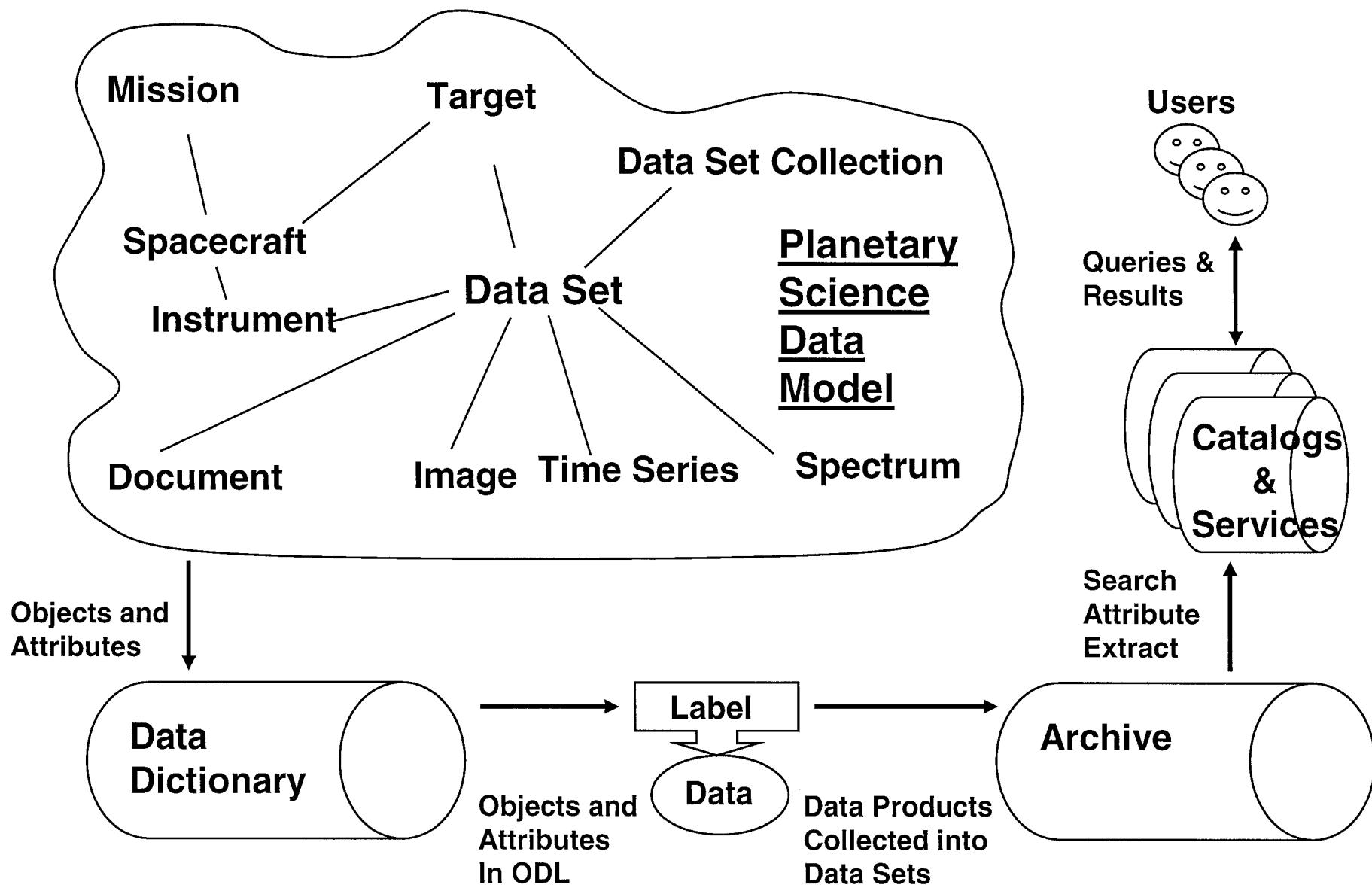
U. Maryland

Node structure provides focus on key disciplines

PDS Standards



PDS Metadata



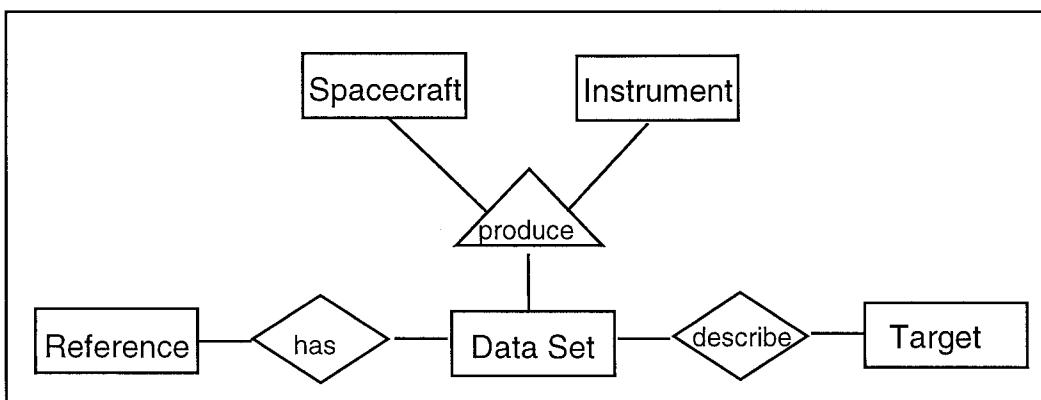
PDS Data Model

Level	Group/Element Structure
1	spacecraft instrument identification group
2	instrument identification
2	instrument name
2	spacecraft identification
2	instrument type
1	instrument description
	...
1	filter group
2	filter name
2	filter number
2	filter type
	...

```

OBJECT          = INSTRUMENT
INSTRUMENT_ID   = VISA
SCID            = VO1
INSTRUMENT_NAME = VISUAL_IMAGING...
INSTRUMENT_TYPE = VIDICON_CAMERA
...
END_OBJECT

```



instinfo				
instid	instname	insttype	scid	...



Language Syntax (PDS Data Product Label - ODL)

DATA_SET_ID	= "VO1/VO2-M-VIS-5-DIM-V1.0"
SPACECRAFT_NAME	= {VIKING_ORBITER_1, ...}
TARGET_NAME	= MARS
IMAGE_ID	= MG88S045
^IMAGE	= 2
SOURCE_IMAGE_ID	= {"383B23", "421B23", ...}
INSTRUMENT_NAME	= {VISUAL_IMAGING_SUBSYSTEM ...}
NOTE	= "MARS DIGITAL IMAGE ..."
OBJECT	= IMAGE
LINES	= 160
LINE_SAMPLES	= 252
SAMPLE_TYPE	= UNSIGNED_INTEGER
SAMPLE_BITS	= 8
SAMPLE_BIT_MASK	= 2#11111111#
CHECKSUM	= 2636242
END_OBJECT	

Planetary Science Data Dictionary (Data Element Definition - ODL)

OBJECT	= ELEMENT_DEFINITION
NAME	= TARGET_NAME
DESCRIPTION	= “The target_name element identifies ...
UNIT	= N/A
VALID_MAXIMUM	= N/A
VALID_MINIMUM	= N/A
MAXIMUM_LENGTH	= 30
MINIMUM_LENGTH	= 1
STANDARD_VALUE_SET	= {MERCURY, VENUS, MARS, ...}
...	
END_OBJECT	= ELEMENT_DEFINITION

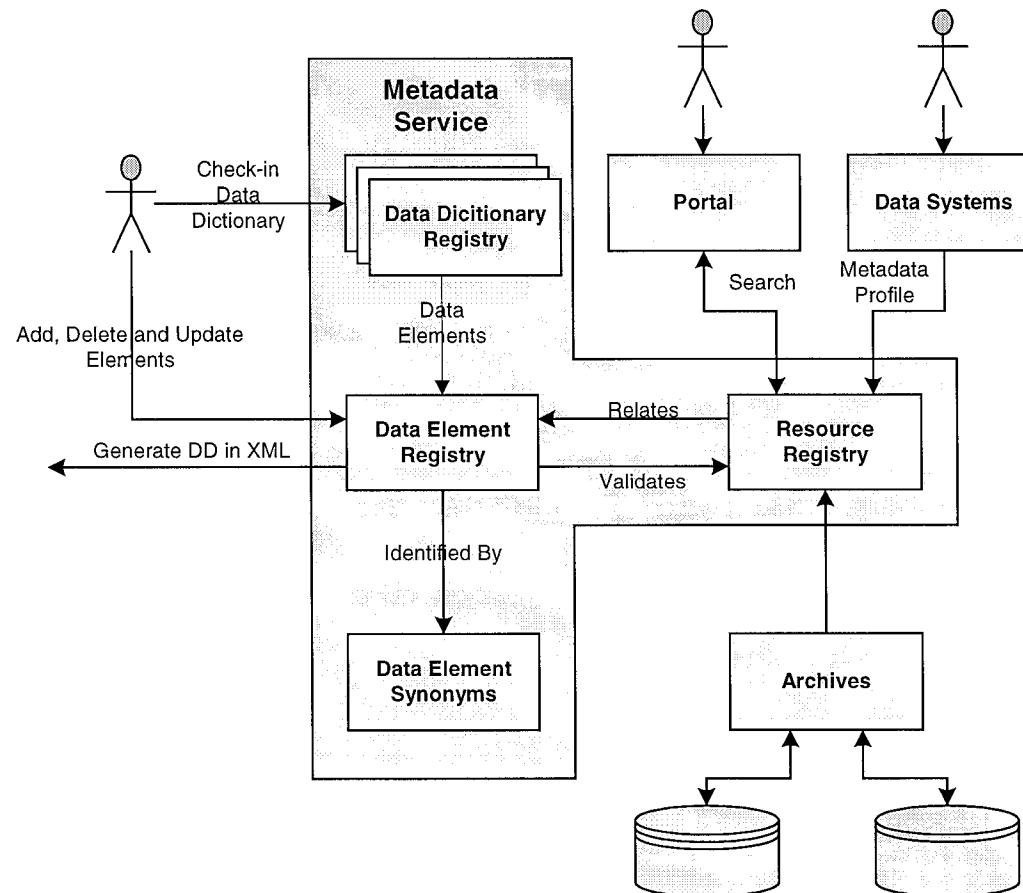
Metadata Service Description

- The purpose of the Metadata Service is to provide a series of registries that allow for the management of data dictionaries, data elements, and resource descriptions.
- The service is designed around three major capabilities:
 - Capturing data dictionary schemas
 - Capturing data elements and schemas
 - Capturing resource definitions for
 - Data systems
 - Data sets
 - Data Products (Images, Documents, etc)

Metadata Service Components

- **Metadata Registry**
 - referred to as a Corporate Metadata Repository (CMR) to manage data dictionary information for projects across enterprise
 - includes a registry of data elements as well as a configuration management system for managing the data dictionary schema
 - data and software engineers can browse the data elements used by each project as well as query the registry through a program interface
- **Resource Registry**
 - manages "profiles" of data system resources across the laboratory based on data elements
 - resources can be definitions of systems, data collections, and data items. Examples of data items include images and documents.

Metadata Service Software Architecture



Metadata Standards

- **Use standards where appropriate**
 - ISO/IEC 11179 – A framework for the Specification and Standardization of Data Elements
 - Dublin Core – A metadata element set intended to facilitate discovery of electronic resources.
 - Emerging protocols for querying and exchanging metadata
 - XMI - XML Metadata Interchange
 - OID – Object Identifier
- **Define laboratory CDEs (Common Data Elements)**
 - CDEs can be discipline, application, or enterprise specific

ISO/IEC 11179

- **Framework for the Specification and Standardization of Data Elements**
- **Is widely accepted specification, provides the underlying basis for data element definition and classification**

ISO/IEC 11179 Implementation

General Descriptors

Descriptor of attribute	Definition of attribute	Obligation	Section in ISO 11179-3
- name	Label assigned to a data element attribute. The name shall be unique and shall be presented as an alphanumeric character string.	M	4.3
- definition	Description of a data element attribute that clearly distinguished it from other data element attributes. The definition is represented as an alphanumeric character string.	M	4.4
- obligation	A descriptor indicating whether a data element attribute shall always be present or sometimes be present	M	4.5
- datatype	A descriptor specifying a set of distinct values for representing the attribute values.	M	4.8
- maximum size	A specification of the maximum number of storage units to represent the distinct values of the datatype specified	M	4.9
- comment	Remark concerning the application of the attributes	O	4.10

ISO/IEC 11179 Implementation

Basic Attributes

Attribute category	Name of data element attribute	Obligation	Maximum size (characters)	Section in (11179-3)
Identifying	name	M	80	6.1.1
	identifier	M	80	6.1.2
	version	M	10	6.1.3
	registration authority	M	35	6.1.4
	synonymous name	O	80	6.1.5
	context	M	80	6.1.6
Definitional	- definition	M	4000	6.2.1
Representational	datatype of data element values	M	35	6.4.3
	maximum size of data element values	M	15	6.4.4
	minimum size of data element values	M	15	6.4.5
	permissible data element values	M	80	6.4.7
Administrative	- comments	O	4000	6.5.4

Dublin Core Implementation

- The Dublin Core element set, a content description model, has received wide spread acceptance amongst the electronic information community
- Metadata Registry uses the Dublin Core element set to describe the check-in data dictionaries
- The Dublin Core Element Set

Content	Property	Instantiation
Title	Creator	Date
Subject	Publisher	Format
Description	Contributor	Identifier
Type	Rights	Language
Source		
Relation		
Coverage		

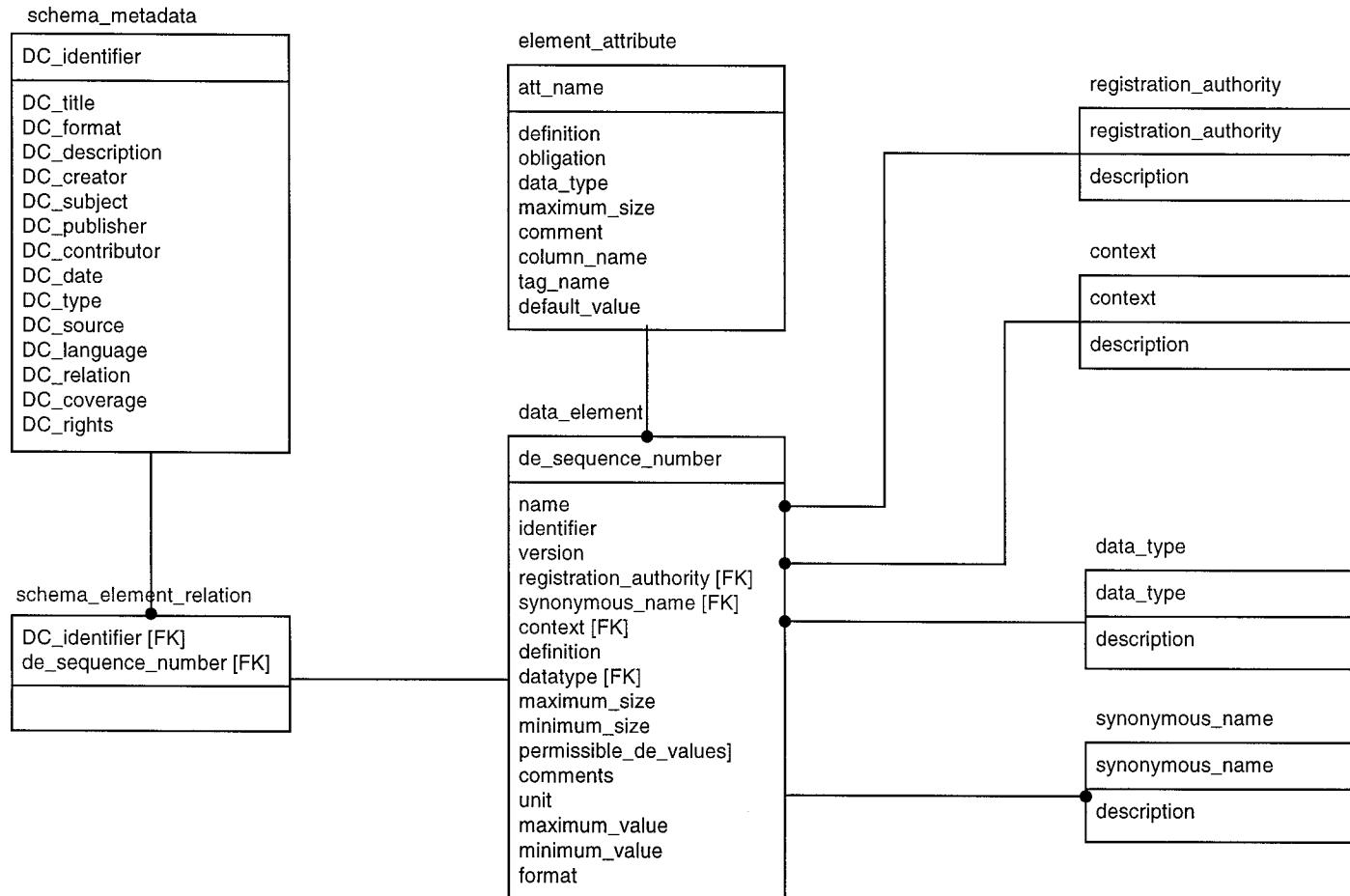
XML & DTD

- **XML (Extensible Markup Language)**
 - Self-describing
 - Combines data and metadata
 - Structured, highly meaningful content
 - XML DTDs describe interchange
- **DTD (XML Document Type Definition)**
 - DTD describes what structures are legal in an XML document
 - They're used by XML processing tools
 - They're plain text, so they're also handy for programmers

Object Identifier (OID)

- Needed to apply to the identifier that it can be persistent, unique, and simple
- **OIDs defined by International Telecommunications Union**
 - Adopted by Internet SNMP and LDAP communities
- **OIDs are persistent that once assigned its lifetime is infinite**
- **OIDs are unique because they follow a distributed management model like the Internet Domain Name System**
- **OIDs are simple because they are a series of short integers**
- **OIDs have the benefit of being able to trace back to its owner**
- **OIDs examples**
 - 1.3.6.1.4.1.1306.1 = EDM
 - 1.3.6.1.4.1.1306.2 = OODT

Metadata Registry Data Model



Metadata Schema DTD

```
<?xml version="1.0" encoding="UTF-8"?>

<!ELEMENT schema (metadata?,dataElement*)>

<!ELEMENT metadata
(DC.Identifier, DC.Title, DC.Format, DC.Description, DC.Creator,
DC.Subject, DC.Publisher, DC.Contributor, DC.Date, DC.Type, DC.Source
DC.Language, DC.Relation, DC.Coverage, DC.Rights)>

<!ELEMENT dataElement
(name, identifier, version, registrationAuthority,
context, definition, dataType, format, unit, minSize, maxSize,
obligation, maxOccurrence, synName? maxValue?, minValue?, comments?)>

<!ELEMENT DC.Identifier (#PCDATA)>
<!ELEMENT DC.Title (#PCDATA)>
<!ELEMENT DC.Format (#PCDATA)>
<!ELEMENT DC.Description (#PCDATA)>
<!ELEMENT DC.Creator (#PCDATA)>
<!ELEMENT DC.Subject (#PCDATA)>
<!ELEMENT DC.Publisher (#PCDATA)>
<!ELEMENT DC.Contributor (#PCDATA)>
<!ELEMENT DC.Date (#PCDATA)>
<!ELEMENT DC.Type (#PCDATA)>
<!ELEMENT DC.Source (#PCDATA)>
<!ELEMENT DC.Language (#PCDATA)>
<!ELEMENT DC.Relation (#PCDATA)>
<!ELEMENT DC.Coverage (#PCDATA)>
<!ELEMENT DC.Rights (#PCDATA)
```

...

PDS Data Element Example

```
<?xml version="1.0"?>
<schema>
...
<dataElement>
    <name>TARGET_NAME</name>
    <identifier>1.3.6.1.4.1.1306.2.10.997</identifier>
    <version>2001</version>
    <registrationAuthority>NASA.PDS</registrationAuthority>
    <context>Metadata.DataDictionary.Element</context>
    <definition>The target_name element identifies a target. The target may be a planet, satellite, ring, region, feature, asteroid or comet. See target_type.</definition>
    <dataType>CHARACTER</dataType>
    <format>N/A</format>
    <unit>none</unit>
    <maxSize>30</maxSize>
    <obligation>Mandatory</obligation>
    <maxOccurrence>1</maxOccurrence>
</dataElement>
...
</schema>
```

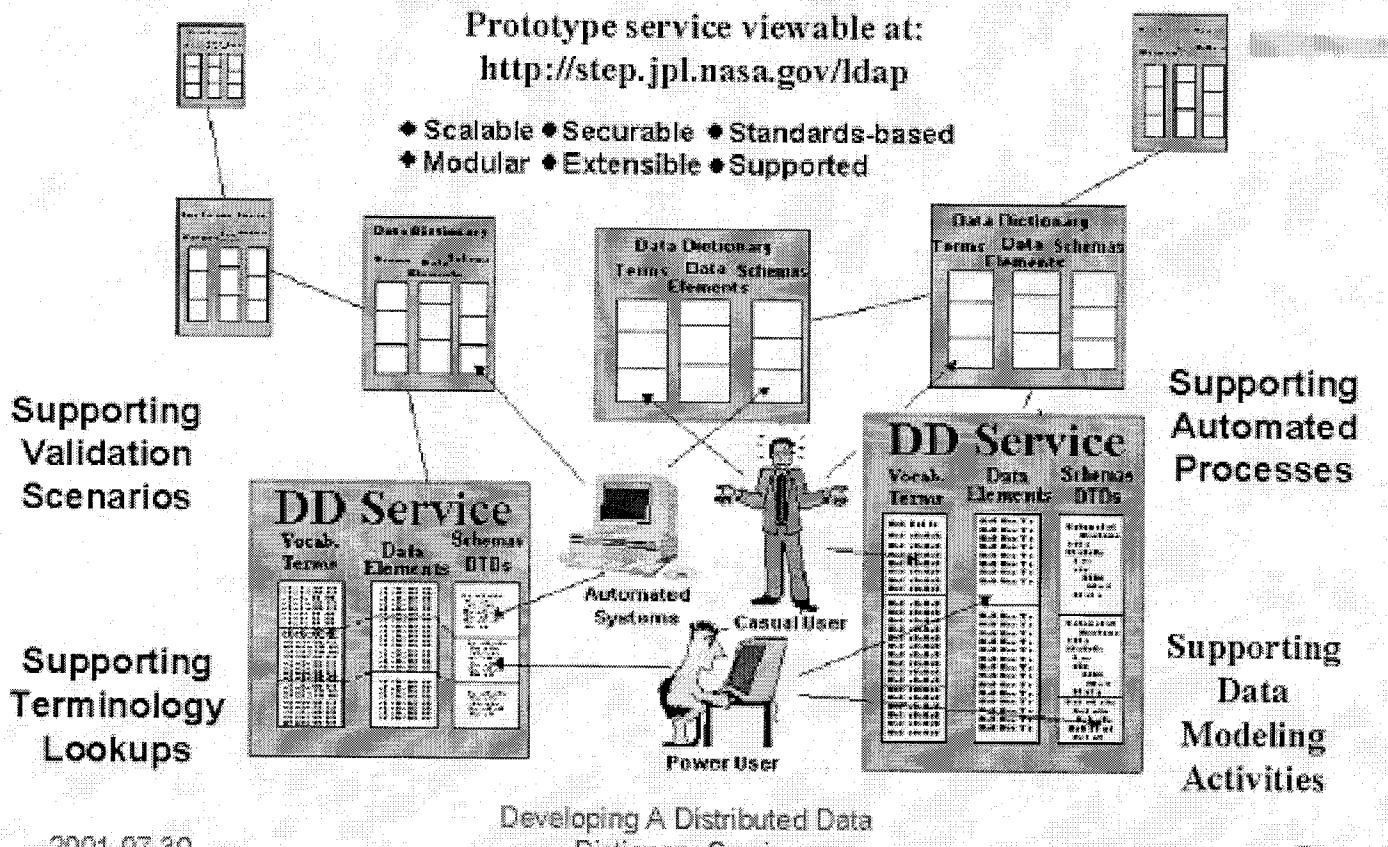
The LDAP Solution

- Develop a distributed data dictionary “service” using:
 - LDAP Internet Protocol (Lightweight Directory Access Protocol)
 - ISO/IEC 11179
 - Dublin Core
 - DSML XML DTD/Schema (Directory Service Markup Language)
- The Service will store and relate vocabulary, data elements, and data model information.

A Distributed Data Dictionary Service

using Standards-based technology

LDAP Protocol | ISO 11179 meta-data schema | DSML | Dublin Core

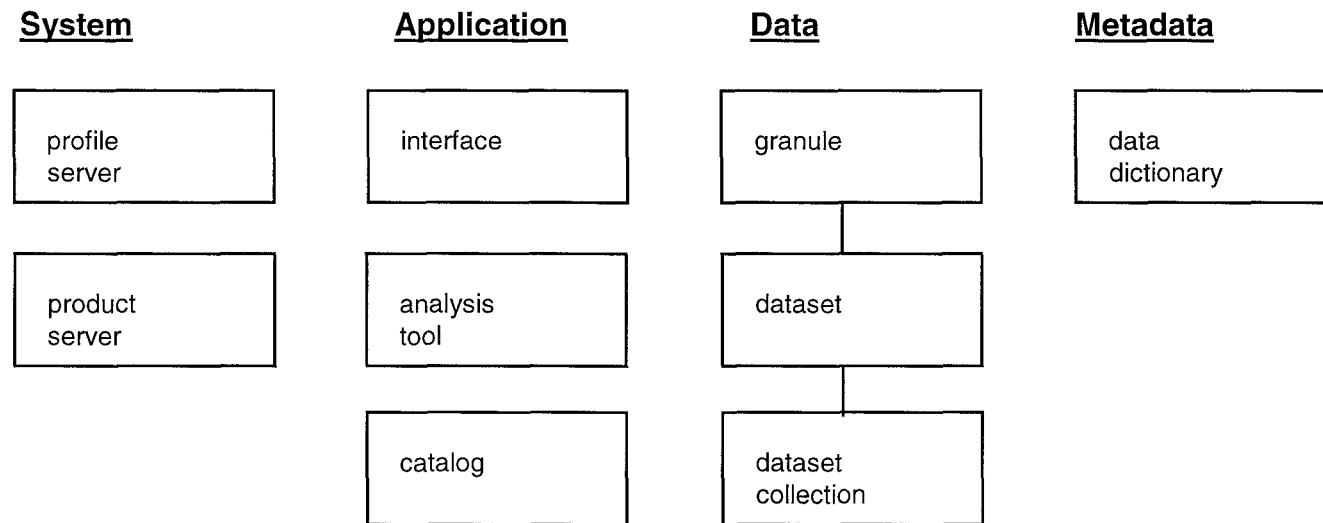


Resource Profile

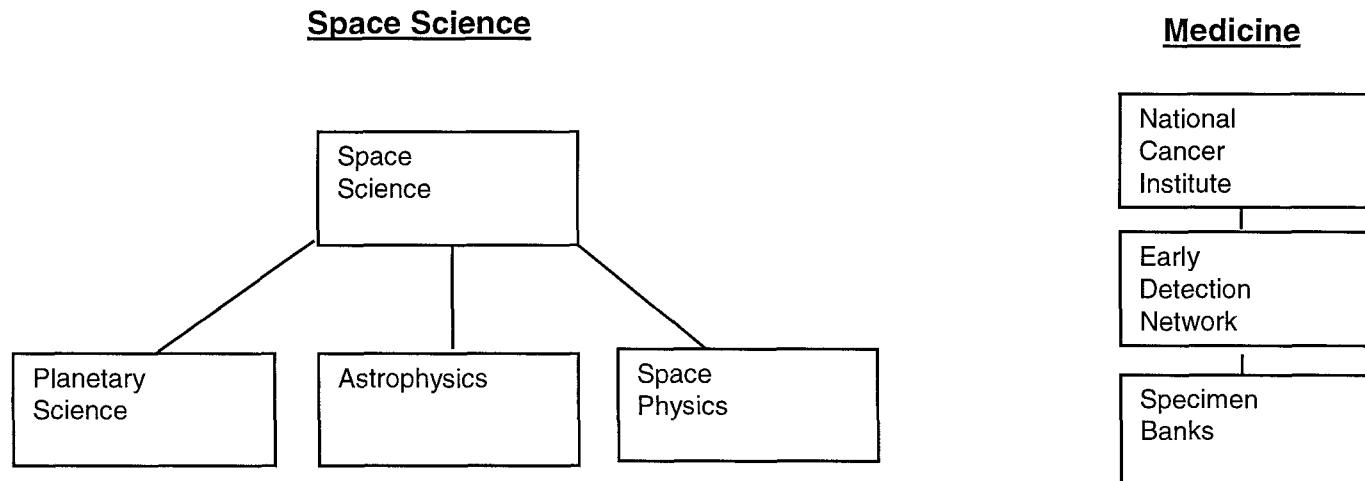
- A profile is a definition of a system resource
 - Consists primarily of metadata
 - Implemented in XML
 - Describes both data and non-data resources
- Managed by profile servers
 - Implemented as Web services in Java
 - Queried using XML based messages
 - Use metadata constraints
 - Implemented as Java classes that map XML profiles to a Java object

Resource Profile Classifications

Resource Classes



Resource Context (Discipline)



Profile DTD

```
<!ELEMENT profiles
  (profile+)>

<!ELEMENT profile
  (profAttributes,
   resAttributes,
   profElement*)>

<!ELEMENT profAttributes
  (profId, profVersion*, profTitle*, profDesc*, profType*,
   profStatusId*, profSecurityType*, profParentId*, profChildId*,
   profRegAuthority*, profRevisionNote*, profDataDictId*)>

<!ELEMENT resAttributes
  (Identifier, Title*, Format*, Description*, Creator*, Subject*,
   Publisher*, Contributor*, Date*, Type*, Source*,
   Language*, Relation*, Coverage*, Rights*,
   resContext*, resAggregation*, resClass*, resLocation*)>

<!ELEMENT profElement
  (elemId*, elemName, elemDesc*, elemType*, elemUnit*,
   elemEnumFlag*, (elemValue | (elemMinValue, elemMaxValue))*/,
   elemSynonym*,
   elemObligation*, elemMaxOccurrence*, elemComment*)>
```

XML Profile Example – Data Set

```
<profile>
  <profAttributes>
    <profId>1.3.6.1.4.1.1306.2.102</profId>
    <profType>profile</profType>
  </profAttributes>
  <resAttributes>
    <Identifier>VO1/VO2-M-VIS-5-DIM-V1.0</Identifier>
    <Title>VO1/VO2_MARS_VISUAL_IMAGING_SUBSYSTEM_DIGITAL...
    <Format>text/html</Format>
    <resContext>NASA.PDS</resContext>
    <resClass>data.dataSet</resClass>
    <resLocation>http://pdsproto.jpl.nasa.gov/catalog/dataset/Resultsds.CFM?...
  </resAttributes>
```

XML Profile Example – Data Set (cont)

```
<profElement>
    <elemName>MISSION_NAME</elemName>
    <elemType>ENUMERATION</elemType>
    <elemValue>VIKING</elemValue>
</profElement>
<profElement>
    <elemName>TARGET_NAME</elemName>
    <elemType>ENUMERATION</elemType>
    <elemValue>MARS</elemValue>
</profElement>
<profElement>
    <elemName>MAXIMUM_LATITUDE</elemName>
    <elemType>REAL</elemType>
    <elemMinValue>-87.50000</elemMinValue>
    <elemMaxValue>90.00000</elemMaxValue>
</profElement>
```

XML Profile Example – Product Server

```
<profile>
  <profAttributes>
    <profId>1.3.6.1.4.1.1306.2.54</profId>
    <profDataDictId>1.3.6.1.4.1.1306.2.10</profDataDictId>
  </profAttributes>
  <resAttributes>
    <Identifier>PDS Product Server</Identifier>
    <Title>PDS Product Server
    <Format>image/pds</Format>
    <Language>en</Language>
    <resContext>PDS</resContext>
    <resAggregation>data.granule</resAggregation>
    <resClass>system.productServer</resClass>
    <resLocation>iiop://JPL.PDS.Product_Server</resLocation>
  </resAttributes>
```

Future work

- Identify, capture, and ingest domain data dictionaries
- Identify common/similar data elements
- Continue to implement ISO/IEC 11179 specifications
- Research MOF (Meta Object Facility)
- Monitor ISO/IEC 11179 implementations
- Integration metadata registry into an enterprise wide information architecture
- Establish best practices for maintaining metadata registry in a operational environment

Other OODT Efforts

- Early Detection Research Network from the National Cancer Institute (NCI)
 - Pilot to federate together epidemiological and biospecimen data from two biomedical cancer research centers
- Children's Hospital, Los Angeles and Johns Hopkins Medical Institute
 - Interested in using JPL OODT technology to link pediatric physiological data between the hospitals
- JPL Enterprise Data Architecture (EDA) effort to build core components to support enterprise-wide data access, data exchange and data management

More Information and References

- **OODT Papers (<http://oodt.jpl.nasa.gov/doc/papers>)**
 - “Science Search and Retrieval using XML” by OODT Team. Presented at Second National Conference on Scientific and Technical Data, National Academy of Sciences, Washington D.C. March 2000.
 - “A Distributed Component Framework for Science Data Product Interoperability” by OODT Team. Presented at the 17th Annual International CODATA conference. Baveno, Italy. October 2000.
- **OODT Presentations (<http://oodt.jpl.nasa.gov/doc/presentations>)**
- **JPL Enterprise Data Architecture White Paper (e-mail: Dan.Crichton@jpl.nasa.gov)**
- **Planetary Data System**
<http://pds.jpl.nasa.gov>
- **Dublin Core**
<http://purl.oclc.org/dc>
- **Extensible Markup Language**
<http://www.w3c.org/XML>
- **ISO/IEC 11179: Specification and Standardization of Data Elements**
- **Federal CIO Statement on Metadata**
<http://www.cio.gov/docs/metadata.htm>